

# Exponential Tracking Error Convergence Bounds for Overparametrized Adaptive Feedforward Systems

David S. Bayard<sup>1</sup>

Jet Propulsion Laboratory  
California Institute of Technology  
4800 Oak Grove Drive  
Pasadena, CA 91109

## Abstract

Many practical adaptive feedforward systems are overparametrized and for this reason will not satisfy persistent excitation (PE) conditions. For these systems, a weaker PE condition is proposed under which it is shown that the tracking error (the error between the desired and estimated outputs) converges exponentially. Bounds are given on the exponential rate of convergence useful for systematic optimization and design purposes. Interestingly, it is also shown that some (but not all) of the robustness properties associated with full PE are retained under the weakened PE condition,

## 1 Introduction

In 1980, Bitmead and Anderson [6] proved that parameter convergence is *exponential* when persistent excitation (PE) conditions are satisfied in the adaptive gradient algorithm. Persistent excitation (PE) conditions are also important for ensuring robustness in a large class of adaptive systems.

Unfortunately, many practical adaptive systems are overparametrized and cannot satisfy a full PE condition. Hence it is of interest to examine more closely what can be attained with only limited excitation. Rather than consider a full PE condition (defined by positive definiteness of the auto correlation matrix), this paper considers a weakened PE condition defined by positive definiteness of a certain “confluence” matrix.

Compared to full PE, the weakened PE condition leads to several interesting consequences:

1. the adaptive system is input-output identical to a reparametrized adaptive system which is *full PE*, but in a smaller number of parameters.

---

<sup>1</sup>Mail Stop 198-326, bayard@bert2.jpl.nasa.gov, Tel: (818) 354-8208, FAX: (818) 393-3444

2. it recovers exponential convergence of the tracking error, but not the parameter error. The tracking error convergence rate is degraded by the condition number of the confluence matrix, analogous to results for the full PE case where exponential rate degrades with the condition number of the autocorrelation matrix.
3. it retains some (but not all) of the robustness properties associated with the full PE case, such as boundedness to bounded disturbances.

These results extend an earlier paper [2] which required that the regressor be periodic. The new confluence matrix condition is much weaker, requiring only that regressor is linearly related to a PE regressor of lower dimension.

The notion of having exponential tracking error convergence without parameter convergence is somewhat unusual, but not completely without precedence in the literature. Using an approximate linear analysis, Glover [7] indicated as early as 1977 that exponential convergence of the tracking error is possible in the adaptive gradient algorithm with an overparametrized tap delay line regressor, and sinusoidal excitation, without any conditions on parameter convergence. More recently, Johansson [9] used a complete end-to-end Lyapunov analysis to demonstrate exponential tracking error convergence (to a bounded set) for a model reference adaptive control (MRAC) algorithm without persistent excitation or parameter convergence. The present research is motivated by these earlier efforts, and in principle can be interpreted as an extension of Glover's work to a much broader class of regressors and adaptive feedforward algorithms.

A brief background is given in Section 2, where the confluence matrix is defined. The main results given in Section 3 show that if the confluence matrix is positive definite the adaptive feedforward operator  $\mathcal{H}$  from error  $e$  to estimate  $\hat{y}$  is input-output identical to an adaptive system with a reduced PE regressor. This is shown in Section 4 to imply that tracking error convergence will be exponential for a large class of overparametrized adaptive feedforward systems. Some analysis of robustness is given in Section 5 and conclusions are postponed until Section 6.

## 2 Background

### 2.1 Adaptive Feedforward Systems

An estimate  $\hat{y}$  of some signal  $y$  is to be constructed as a linear combination of the elements of a regressor vector  $x(t) \in RN$ , i.e.,

*Estimated Signal*

$$\hat{y} = w(t)^T x(t) \tag{2.1}$$

where  $w(t) \in RN$  is a parameter vector which is tuned in real-time using the adaptation algorithm,

### Adaptation Algorithm

$$w = \mu \Gamma(p)[\tilde{x}(t)e(t)] \quad (2.2)$$

Here,  $\Gamma(p)[\cdot]$  denotes the multivariable LTI transfer function  $\Gamma(s) \cdot I$  where  $\Gamma(s)$  is any SISO LTI transfer function in the Laplace  $s$  operator (the differentiation operator  $p$  will replace the Laplace operator  $s$  in all time-domain filtering expressions); the term  $e(t) \in \mathbb{R}^1$  is an error signal;  $\mu > 0$  is an adaptation gain; and the signal  $\tilde{x}$  is obtained by filtering the regressor  $x$  through any stable filter  $F(p)$ , i.e.,

#### Regressor Filtering

$$\tilde{x} = F(p)[x] \quad (2.3)$$

The notation  $F(p)[\cdot]$  denotes the multivariable LTI transfer function  $F(s) \cdot I$  with SISO filter  $F(s)$ , acting on the indicated vector time domain signal.

Equations (2.1)-(2.3) taken together will be referred to as an *adaptive feedforward system*. Collectively, these equations define an important open-loop mapping from the error signal  $e$  to the estimated output  $\hat{y}$ . Because of its importance, the mapping from  $e$  to  $\hat{y}$  will be denoted by the special character  $\mathcal{H}$ , i.e.,

$$\hat{y} = \mathcal{H}[e] \quad (2.4)$$

The special structure of  $\mathcal{H}$  is depicted in Figure 2.1.

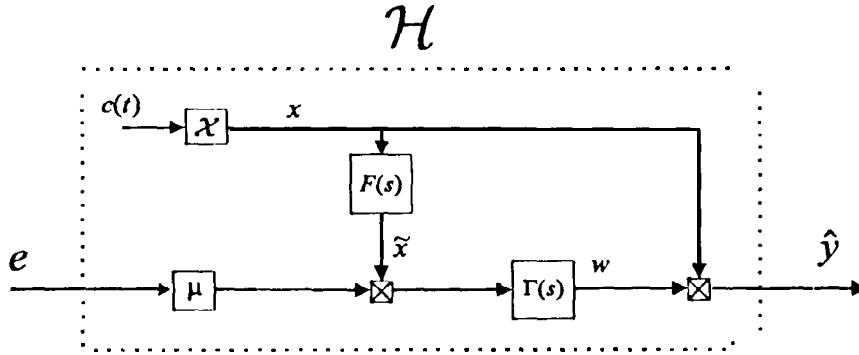


Figure 2.1: LTV operator  $\hat{y} = \mathcal{H}[e]$  for adaptive system with regressor  $\tilde{x}$ , adaptation law  $\Gamma(s)$ , and regressor filter  $F(s)$

**REMARK 2.1** The definition of  $\Gamma(s)$  is left intentionally general to include analysis of the gradient algorithm (i.e., with the choice  $\Gamma(s) = 1/s$ ), the gradient algorithm with leakage (i.e.,  $\Gamma(s) = 1/(s + \sigma)$ ;  $\sigma \geq 0$ ), proportional-plus-integral adaptation (i.e.,  $\Gamma(s) = k_p + k_i/s$ ), or arbitrary linear adaptation algorithms of the designer's choosing. Adaptation laws which are nonlinear or normalized (e.g., divided by the norm of the regressor), are not considered here since they do not have an equivalent LTI representation  $\Gamma(s)$ . ■

**REMARK 2.2** The use of the regressor filter  $F(s)$  in (2.3) allows the unified treatment of many important adaptation algorithms including the well-known Filtered-X algorithm from the signal processing literature [13], and the Augmented Error algorithm of Monopoli (cf., [10]). ■

## 2.2 Confluence Matrices and Overparametrization

Let  $c(t) \in R^n$  be a bounded piecewise continuous signal vector, and let there exist positive constants  $\beta_1, \beta_2, T_o > 0$  such that,

$$\beta_1 \cdot I \leq \int_t^{t+T_o} c(\tau)c^T(\tau)d\tau \leq \beta_2 \cdot I \quad (2.5)$$

for all  $t \geq 0$ . Any signal  $c(t)$  which satisfies these properties is said to be Persistently Exciting (PE) with bounds  $\{\beta_1, \beta_2, T_o\}$  [8].

For the purpose of this paper, it will be assumed that the regressor  $x(t) \in R^N$  is linearly related to such a PE signal  $c(t)$  as follows,

$$x = Xc(t) \quad (2.6)$$

where  $X \in R^{N \times n}$ . It is also assumed that  $N \geq n$  in (2.6), so that  $X$  is a “tall” matrix and the adaptive feedforward system is *overparametrized*.

**DEFINITION 2.1** The matrix  $X^T X$  is defined as the **confluence matrix** associated with a particular regressor  $x$  of the form (2.6). m

The name “confluence matrix” has been chosen to reflect the fact that  $N$  signal channels seen at the output of the tall matrix  $X$  are effectively combined into a smaller number of  $n$  channels ( $n \leq N$ ) when forming  $X^T X$ . The confluence matrix will play an important role in characterizing the convergence properties of the overparametrized adaptive system.

Unfortunately, if the system is overparametrized the PE condition is impossible to satisfy. This will be shown by example.

**EXAMPLE 2.1** For simplicity, consider the case where  $c$  is periodic, i.e.,

$$c(t) = [\sin \omega_o t, \cos \omega_o t, \sin 2\omega_o t, \cos 2\omega_o t, \dots, \sin(m\omega_o t), \cos(m\omega_o t)]^T \quad (2.7)$$

Letting  $T_o = 2\pi/\omega_o$  and using (2.7) one can calculate,

$$\int_t^{t+T_o} c(\tau)c^T(\tau)d\tau = \frac{T_o}{2} \cdot I \in R^{n \times n} \quad (2.8)$$

Hence, the PE condition (2.5) is satisfied with  $\beta_1 = \beta_2 = T_o/2$ .

Because of the form of  $c(t)$  in (2.7), any regressor  $x = \mathcal{X}c(t)$  will be periodic with period  $T_o = 2\pi/\omega_o$ . One can check the PE condition by computing the *autocorrelation matrix*,

$$\int_t^{t+T_o} x(\tau)x^T(\tau)d\tau = \mathcal{X} \int_t^{t+T_o} c(\tau)c^T(\tau)d\tau \mathcal{X}^T \quad (2.9)$$

$$= \frac{T_o}{2} \mathcal{X} \mathcal{X}^T \quad (2.10)$$

It is seen that the autocorrelation matrix is essentially the *outer-product* of the matrix  $\mathcal{X}$  with itself. Consequently, if the problem is overparametrized (i.e.,  $N > n$ ) the matrix  $\mathcal{X}$  is “tall”, and it is impossible for  $\mathcal{X}\mathcal{X}^T$  to be positive definite i.e., it is impossible to satisfy the PE conditions. m

It is a common belief that along with the loss of PE, comes the loss of exponential convergence. The main point of this paper is to show that this is not generally true, and in fact the exponential convergence properties of the tracking error  $e$  (and the parameters on a reduced subspace) do not depend on the *outer product condition*  $\mathcal{X}\mathcal{X}^T > 0$ , but rather depend on the *inner product condition*  $X^T X > 0$ .

### 3 Regressor Reduction to PE

The next result shows that if the confluence matrix is positive definite, the adaptive feed-forward operator  $\mathcal{H}$  can always be reparametrized to have a PE regressor without changing its input-output properties.

**THEOREM 3.1 (Regressor Reduction to PE)** *Let the confluence matrix associated with the adaptive feedforward system (2.1)-(2.3) be positive definite,*

$$X^T X > 0 \quad (3.1)$$

Then,

(i) *The input-output properties of the LTV operator  $\mathcal{H}$  from  $e$  to  $\hat{y}$  are invariant under the change of variables,*

$$\eta(t) = \Lambda^{-\frac{1}{2}} P \mathcal{X}^T x(t) \quad (3.2)$$

$$\rho(t) = \Lambda^{-\frac{1}{2}} P \mathcal{X}^T w(t) \quad (3.3)$$

Here,  $\eta \in R^n$  and  $p \in R^*$  are reduced-order regressor and parameter vectors, respectively, and matrices  $P, \Lambda \in R^{n \times n}$  are defined from the eigenvalue decomposition of the confluence matrix,

$$\mathcal{X}^T \mathcal{X} = P^T \Lambda P \quad (3.4)$$

$$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\} > 0 \quad (3.5)$$

where  $P^T = P^{-1}$ , and it is assumed that the eigenvalues are ordered as  $\lambda_1 \geq \dots \geq \lambda_n > 0$ .

(ii) The reduced-order regressor  $\eta \in R^n$  is PE with the bounds,

$$\beta_1 \lambda_n \cdot I \leq \int_t^{t+T_o} \eta(\tau) \eta^T(\tau) d\tau \leq \beta_2 \lambda_1 \cdot I \quad (3.6)$$

for all  $t \geq 0$ , where  $\beta_1, \beta_2, T_o$  are defined by the PE condition (2.5) for  $c(t)$ .

**PROOF:**

**Proof of (i)** The proof follows simply by the superposition and scaling properties of linear operators. As such, it can be proved graphically. Consider the sequence of block diagram rearrangements shown in Figure 3.1. Specifically, Figure 3.1 Part a. shows the initial adaptive system with overparametrized regressor  $x$ ; Part b. shows the matrix  $\mathcal{X}$  pushed through several scalar matrix blocks of the diagram; Part c. replaces the confluence matrix by its eigenvalue decomposition  $\mathcal{X}^T \mathcal{X} = P^T A P$ ; Part d. pushes the matrix factor  $\Lambda^{\frac{1}{2}} P$  back through several scalar matrix blocks. The resulting block diagram is driven by the regressor  $\eta$  which related to  $c$  by the nonsingular transformation  $\eta = \Lambda^{\frac{1}{2}} P c$ , and hence is PE.

**Proof of (ii)** Define,

$$M = \int_t^{t+T_o} \eta(\tau) \eta^T(\tau) d\tau \quad (3.7)$$

Substituting  $\eta = \Lambda^{\frac{1}{2}} P c$  into (3.7), gives

$$M = \Lambda^{\frac{1}{2}} P \int_t^{t+T_o} c(\tau) c^T(\tau) d\tau P^T \Lambda^{\frac{1}{2}} \quad (3.8)$$

Using the PE property of  $c(t)$  in (2.5) and (3.5) gives,

$$\bar{\sigma}(M) \leq \beta_2 \lambda_1 \quad (3.9)$$

$$\underline{\sigma}(M) \geq \beta_1 \lambda_n \quad (3.10)$$

which is equivalent to (3.6) as desired. ■

In order to use the result of Theorem 3.1, the overparametrized regressor  $x \in RN$  must satisfy the decomposition  $x(t) = \mathcal{X}_1 c_1(t)$  for some matrix  $\mathcal{X}_1 \in R^{N \times n}$ , and some PE signal vector  $c_1 \in R^n$ . Fortunately, in many applications such a decomposition is straightforward to find. Consider the next example.

**EXAMPLE 3.1** [Harmonic Regressor] It is common for the regressor  $x$  to have a harmonic form, i.e., its elements are composed of linear combinations of  $m$  sinusoidal frequencies  $\{\omega_i\}_{i=1}^m$ . For this regressor, the signal  $c_1$  can be chosen of length  $2m$  having elements,

$$c_1(t) = [\sin \omega_1 t, \cos \omega_1 t, \dots, \sin \omega_m t, \cos \omega_m t]^T \in R^{2m} \quad (3.11)$$

This construction guarantees the existence of a matrix  $\mathcal{X}_1 \in R^{2m}$  in the desired decomposition  $x = \mathcal{X}_1 c_1(t)$ . ■

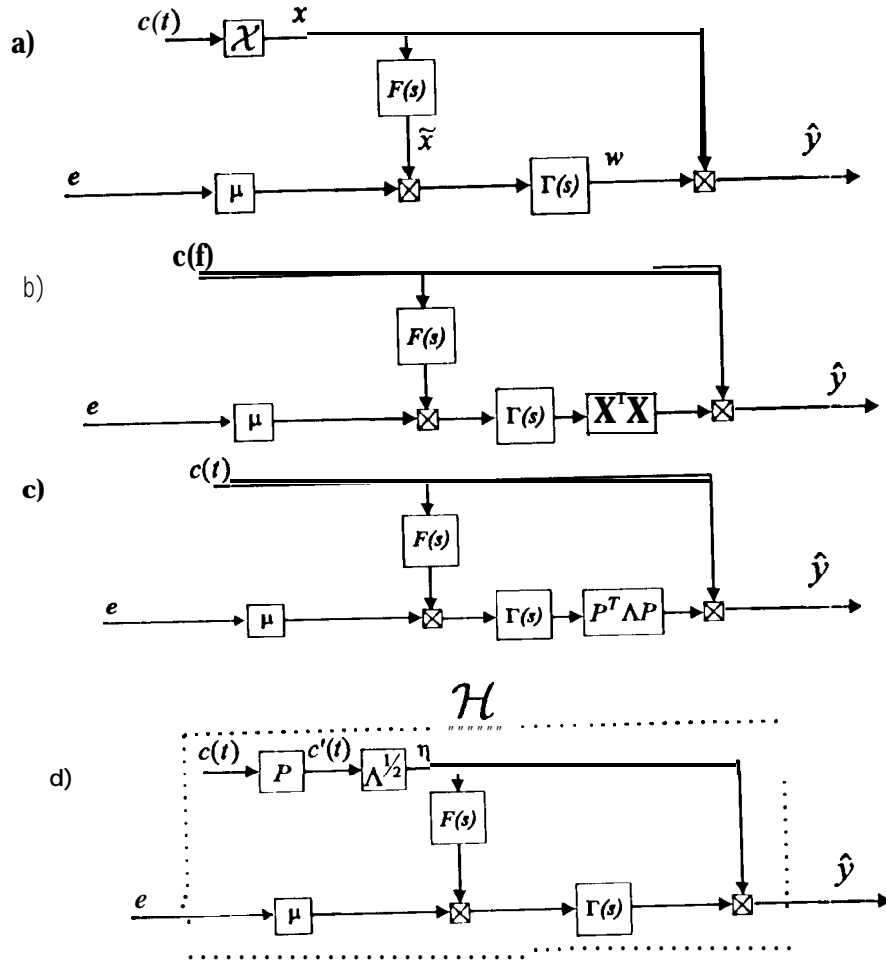


Figure 3.1: Proof of Theorem 3.1 by Block diagram prearrangements

**REMARK 3.1** When the regressor has the harmonic form  $x = \mathcal{X}c$  for  $c$  of the form (3.11), it sometimes turns out that the mapping  $\mathcal{H}$  is *purely Linear Time-Invariant (LTI)*. This LTI property considerably simplifies the analysis, and lies at the heart of Glover's approach in [7]. It has been found recently [4] that this LTI property occurs *if and only if the confluence matrix has a pairwise diagonal structure*, i.e.,

$$\mathcal{X}^T \mathcal{X} = D^2 \quad (3.12)$$

where,

$$D^2 = \text{diag}[d_1^2, d_1^2, d_2^2, d_2^2, \dots, d_m^2, d_m^2] \quad (3.13)$$

■

The input-output properties of  $\mathcal{H}$  play a critical role in determining the convergence properties of the tracking error  $e$  in closed-loop. The result of Theorem 3.1 is important because it shows that the input-output properties of the operator  $\mathcal{H}$  (even overparametrized) are identical to one which is reparametrized to have a (reduced-order) PE regressor. Given such a PE regressor, many proofs of exponential stability exist in the literature (cf., [8] [10] [11]). As applied to the present case, these proofs ensure exponential convergence of the tracking error (and the parameters on a reduced subspace) in closed-loop. The adaptive gradient algorithm is chosen in the next section as a simple and representative algorithm to examine these exponential convergence properties in more detail.

## 4 Exponential Convergence

### 4.1 Adaptive Gradient Algorithm

Let the  $y(t) \in R^1$  and  $x(t) \in R^N$ , be known signals and assume there exists a constant parameter vector  $w^o \in RN$  such that,

$$y(t) = w^{oT} x(t) \quad (4.1)$$

for all  $t > 0$ . Uniqueness of  $w^o$  is not required (i.e., the system can be overparametrized). An estimate  $\hat{y}$  of  $y$  is constructed as,

$$\hat{y} = w(t)^T x(t) \quad (4.2)$$

where  $w(t)$  is tuned in real-time using the adaptive gradient algorithm [10] (i.e., set  $\Gamma(s) = 1/s$  in (2.2) and  $F(s) = 1$  in (2.3) ),

$$w = \mu x(t) e(t) \quad (4.3)$$

with adaptation gain  $\mu > 0$ . The tracking error is defined as,

$$e(t) = y(t) - \hat{y}(t) \quad (4.4)$$



and the parameter error is defined as,

$$\phi(t) = w^o - w(t) \quad (4.5)$$

Using (4.1)(4.2)(4.4)(4.5), the tracking and parameter errors can be related as follows,

$$\mathbf{e} = \phi^T x(t) \quad (4.6)$$

Assuming that the true parameter  $w^o$  does not vary with time, (i.e.,  $\dot{w}^o = 0$ ), it follows from (4.3)(4.5) that,

$$\dot{\phi} = \dot{w}^o - \dot{w} = -\mu x e = -\mu x x^T \phi \quad (4.7)$$

This equation characterizes the propagation of the parameter error.

## 4.2 Exponential Convergence Properties

It is convenient at this point to review a well-known stability argument. Define the Lyapunov function candidate,

$$V = \frac{1}{2} \phi^T \phi \quad (4.8)$$

Taking the derivative of (4.8) and using (4.1)-(4.7) yields,

$$\dot{V} = -\mu e \phi^T x = -\mu e^2 \leq 0 \quad (4.9)$$

This proves that  $\phi$  remains bounded. If  $x$  is bounded, then from (4.6) the error  $e$  remains bounded. Furthermore, if  $\dot{x}$  is bounded, then  $\dot{V}$  is bounded,  $V$  is uniformly continuous, and Barbalat's lemma ([10], pg. 85, and 276), can be applied to ensure that  $\lim_{t \rightarrow \infty} e = 0$ . This well known argument ensures that the error converges to zero as desired.

While the above argument ensures that  $e$  converges to zero, it does not indicate *how fast* it converges. Additional conditions such as persistent excitation are typically imposed which ensure exponential convergence of  $e$  to zero.

Persistent excitation conditions in adaptive algorithms have been studied by many researchers. Early results can be found in Astrom and Bohlin [1] where the PE condition is expressed in terms of positive definiteness of the autocorrelation function formed from the regressor. Subsequently, Bitmead and Anderson [6] proved that parameter convergence is *exponential* when PE conditions are satisfied in the adaptive gradient algorithm and the normalized adaptive gradient algorithms. Explicit upper and lower bounds on the exponential response can be found in [12]. A general discussion of the PE condition is given in [5] and an effort to unify many definitions can be found in [14].

As an example, consider the case without overparametrization (i.e.,  $N = n$ ), so that  $x(t) \in R^n$  is bounded and PE satisfying, say,

$$\|x(t)\| \leq \bar{x} < \infty; \text{ for all } t \geq 0 \quad (4.10)$$

$$\alpha_1 \cdot I \leq \int_t^{t+\delta} x(\tau)x^T(\tau)d\tau \leq \alpha_2 \cdot I \quad (4.11)$$

for some  $\alpha_1, \alpha_2, \delta > 0$ . Then it is well known (cf., [10][11] [8]), that the error  $e$  converges exponentially. Specifically, there exist constants  $p \geq 0, \alpha > 0$  such that,

$$|e| \leq p e^{-\alpha m} \quad (4.12)$$

The precise expression for  $\alpha$  is given in Lemma A.1 of Appendix A as,

$$\alpha = \frac{1}{2\delta} \ln \left( \frac{1}{1 - \alpha_3} \right) \quad (4.13)$$

$$\alpha_3 = \frac{1}{(1 - \alpha_3)^{\frac{1}{2}}} \|\phi(0)\| \quad (4.14)$$

$$\alpha_3 = \frac{2\mu\alpha_1}{(1 + \mu\alpha_2\sqrt{n})^2} \quad (4.15)$$

The convergence rate  $\alpha$  in (4.13) is a function of  $\mu$  through the expression (4.15). For small  $\mu$  the rate can be approximated by,

$$\alpha \simeq \frac{\mu\alpha_1}{\delta} \quad (4.16)$$

The fastest convergence rate is found by optimizing  $\alpha$  in (4.13) with respect to  $\mu$ . Specifically, the condition  $d\alpha/d\mu = 0$  can be solved to give the optimal gain as,

$$\mu^* = \frac{1}{\alpha_2\sqrt{n}} \quad (4.17)$$

Substituting (4.17) into (4.13) gives,

$$\alpha^* = \frac{1}{2\delta} \ln \left( \frac{1}{1 - \frac{1}{2\sqrt{n}} \cdot \frac{\alpha_1}{\alpha_2}} \right) \quad (4.18)$$

It is seen that the optimized rate  $\alpha^*$  improves monotonically with the ratio  $\alpha_1/\alpha_2$ . This ratio is precisely the reciprocal condition number of the autocorrelation matrix (4.11), and motivates keeping this condition number as close to unity as possible for fast convergence (assuming it is optimally tuned with  $\mu^*$ ).

Exponential convergence for the overparametrized adaptive gradient algorithm is examined next.

**THEOREM 4.1 (Overparametrized Adaptive Gradient)** Assume there exists a  $w \in \mathbb{R}^N$  such that (4.1) holds for all  $t \geq 0$ , and that the adaptive gradient algorithm (4.2)-(4.7) is used to tune  $w$ , giving the following error system,

$$e = \phi^T x \quad (4.19)$$

$$\dot{\phi} = -\mu x x^T \phi \quad (4.20)$$

Let  $c(t) \in \mathbb{R}^n$  be a bounded piecewise continuous signal vector which is PE, i.e., let there exist positive constants  $\beta_1, \beta_2, T_0 > 0$  such that,

$$\beta_1 \cdot I \leq \int_t^{t+T_0} c(\tau) c(\tau)^T d\tau \leq \beta_2 \cdot I; \text{ for all } t \geq 0 \quad (4.21)$$

$$\|c(t)\| \leq \bar{c} < \infty; \text{ for all } t \geq 0 \quad (4.22)$$

Let the regressor  $x(t) \in \mathbb{R}^N$  be linearly related to the PE signal  $c(t)$  as follows,

$$x = \mathcal{X}c(t) \quad (4.23)$$

where  $\mathcal{X} \in \mathbb{R}^{N \times n}$  and  $N \geq n$  (i. e., the system can be overparametrized). Let the confluence matrix be positive definite,

$$\mathcal{X}^T \mathcal{X} > 0 \quad (4.24)$$

and let the eigenvalue decomposition of the confluence matrix be given as,

$$\mathcal{X}^T \mathcal{X} = P^T \Lambda P \quad (4.25)$$

$$\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\} > 0 \quad (4.26)$$

where  $P^T = P^{-1}$ , and  $\lambda_1 \geq \dots \geq \lambda_n > 0$ .

Then,

(i) The error system (4.19)(4.20) can be written equivalently as the reduced system,

$$e = r^T \eta \quad (4.27)$$

$$\dot{r} = -\mu \eta \eta^T r \quad (4.28)$$

where the reduced regressor  $\eta \in \mathbb{R}^n$  and parameter error  $r \in \mathbb{R}^n$  are given, respectively, by,

$$\eta = \Lambda^{\frac{1}{2}} P c(t) \quad (4.29)$$

$$r = \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \phi \quad (4.30)$$

(ii) The reduced regressor  $\eta \in \mathbb{R}^n$  is PE with the bounds,

$$\beta_1 \lambda_n \cdot I \leq \int_t^{t+T_0} \eta(\tau) \eta^T(\tau) d\tau \leq \beta_2 \lambda_1 \cdot I \quad (4.31)$$

for all  $t \geq 0$ .

(iii) The tracking error  $e$  and reduced parameter error  $r$  converge to zero exponentially as,

$$\|r\| \leq p_0 e^{-\alpha t} \quad (4.32)$$

$$|e| \leq p_0 \lambda_1^{\frac{1}{2}} \bar{c} e^{-\alpha t} \quad (4.33)$$

where,

$$\alpha = \frac{1}{2T_o} \ln \left( \frac{1}{1 - \alpha_3} \right) \quad (4.34)$$

$$p_0 = \left( \frac{1}{1 - \alpha_3} \right)^{\frac{1}{2}} \|r(0)\| \quad (4.35)$$

$$\alpha_3 = \frac{2\mu\beta_1\lambda_n}{(1 + \mu\beta_2\lambda_1\sqrt{n})^2} \quad (4.36)$$

Letting  $\mu$  be sufficiently small (i. e., such that  $\mu \ll 1/(\beta_2\lambda_1T_o\sqrt{n})$ ), gives,

$$\alpha \simeq \mu\beta_1\lambda_n/T_o \quad (4.37)$$

$$p_0 \simeq (1 + \mu\beta_1\lambda_n)\|r(0)\| \quad (4.38)$$

## PROOF:

Proof of (i): Using the transformed vectors  $\eta$  and  $r$ , the error equation (4.19) can be written as follows,

$$e = \phi^T x = \phi^T \mathcal{X}c(t) \quad (4.39)$$

$$= \phi^T \mathcal{X}P^T \Lambda^{-\frac{1}{2}} \Lambda^{\frac{1}{2}} P c(t) \quad (4.40)$$

$$= r^T \eta \quad (4.41)$$

which is (4.27) as desired. Likewise, for the adaptation law (4.20) one has,

$$\dot{\phi} = -\mu x x^T \phi = -\mu \mathcal{X} c c^T \mathcal{X}^T \phi \quad (4.42)$$

Multiplying both sides of (4.42) on the left by  $\Lambda^{-\frac{1}{2}} P \mathcal{X}^T$  gives,

$$\Lambda^{-\frac{1}{2}} P \mathcal{X}^T \dot{\phi} = -\mu \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \mathcal{X} c c^T \mathcal{X}^T \phi \quad (4.43)$$

$$= -\mu \Lambda^{-\frac{1}{2}} P P^T \Lambda P c c^T \mathcal{X}^T \phi \quad (4.44)$$

$$= -\mu \Lambda^{\frac{1}{2}} P c c^T \mathcal{X}^T \phi \quad (4.45)$$

$$= -\mu \Lambda^{\frac{1}{2}} P c c^T P^T \Lambda^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} P \mathcal{X}^T \phi \quad (4.46)$$

Substituting (4.29) and (4.30) into both sides of (4.46) gives,

$$\dot{r} = -\mu\eta\eta^T r \quad (4.47)$$

which is (4.28) as desired.

Proof of (ii): Identical to Proof of (ii) of Theorem 3.1.

Proof of (iii) By Part (ii) the reduced regressor  $\eta$  in (4.29) is persistently exciting. It follows from Lemma A. 1 of Appendix A that the reduced error system (4.27)(4.28) converges exponentially. In light of the PE bounds in (4.31), Lemma A.1 can be applied with  $\alpha_1 = \beta_1 \lambda_n$ ,  $\alpha_2 = \beta_2 \lambda_1$ ,  $\delta = T_o$  and  $\bar{x} = \lambda_1^{\frac{1}{2}} \bar{c}$  to give results (4.32)-(4.38) as desired. ■

### 4.3 Discussion

Intuitively, the persistent excitation conditions are eliminated in Theorem 3.1 by avoiding the need for convergence of the full parameter vector  $w$  in the proof. Rather, the “degree” to which the given regressor  $x$  is persistently exciting is indicated by the size  $n$  of the vector  $c(t)$ . The parameter error vector  $\phi$  is transformed to become the smaller vector  $r = A^{-1} P \mathcal{X}^T \phi$  where  $r \in R^n$  is defined on a subspace which is excited persistently. Since the regressor  $\eta$  associated with  $r$  is persistently exciting, the reduced error vector  $r$  converges exponentially, which from (4, 27) ensures exponential convergence of  $e$ .

The exponential rate of convergence will be examined more closely. If one constructs the ideal regressor as  $x(t) = c(t)$  then the regressor is PE with parameters  $\{\beta_1, \beta_2, T_o\}$  and one would achieve an optimal exponential convergence of (set  $(\alpha_1 = \beta_1, \alpha_2 = \beta_2, \delta = T_o$  in (4.18)),

$$\alpha^* = \frac{1}{2\delta} \ln \left( \frac{1}{1 - \frac{1}{2\sqrt{n}} \cdot \frac{\beta_1}{\beta_2}} \right) \quad (4.48)$$

However, if the regressor is overparametrized as  $x = \mathcal{X}c(t)$ , then Theorem 4.1 part (ii) indicates that the PE parameters of the equivalent reduced-order regressor  $\eta$  are degraded to  $\beta_1 \lambda_n, \beta_2 \lambda_1$ , TO, This modifies the optimal convergence rate to,

$$\alpha^* = \frac{1}{2\delta} \ln \left( \frac{1}{1 - \frac{1}{2\sqrt{n}} \cdot \frac{\beta_1}{\beta_2} \cdot \frac{\lambda_n}{\lambda_1}} \right) \quad (4.49)$$

i.e., the rate is degraded by (a monotonic function of) the ratio  $\lambda_n/\lambda_1$ . This ratio is precisely the reciprocal of the condition number of the confluence matrix  $\mathcal{X}^T \mathcal{X}$ , denoted as,

$$\kappa(\mathcal{X}^T \mathcal{X}) \triangleq \frac{\lambda_1}{\lambda_n} \quad (4.50)$$

In words then, *the optimal exponential convergence rate for an overparametrized adaptive system degrades (compared to the full PE case) monotonically with the condition number of the confidence matrix.*

The reader is warned that this condition number can be quite large. *In fact, most of the bad experiences that researchers have with sluggish convergence of overparametrized adaptive systems can be traced to this quantity.* The usual tendency is to blame overparametrization and lack of PE. However, the main point to be made here is that, on the contrary, overparametrized systems have exponential convergence just like full PE systems, and can provide good performance if care is taken to ensure that the condition number (4.50) is well behaved.

## 5 Robustness Properties

Generally speaking, a full PE condition contributes significantly to the overall robustness of the adaptive system due to exponential convergence of the *entire parameter vector*. Unfortunately, in the case of the weakened PE condition the parameters converge exponentially only on a reduced subspace and one cannot expect the same level of robustness. Nevertheless, it is shown that a certain level of robustness is retained, in particular, those robustness properties that depend solely on the input/output properties of the adaptive algorithm.

If a bounded disturbance  $d$  is added to the *output*  $y$  in (4.1), then the error system (4.19)(4.20) becomes,

$$e = \phi^T x + d \quad (5.1)$$

$$\dot{\phi} = -\lambda \phi + \lambda x e - \mu x d \quad (5.2)$$

Using the identities in Theorem 4.1, the reduced error system can be calculated as,

$$e = r^T \eta + d \quad (5.3)$$

$$\dot{r} = -\mu \eta \eta^T r - \mu \eta d \quad (5.4)$$

Since the reduced regressor  $\eta$  is PE, the linear time-varying equation (5.4) is exponentially stable, and it can be shown using standard results (cf., [8], Chapter 9) that the signals  $e$  and  $r$  will remain bounded. Hence the exponential convergence property ensured by the partial PE condition is robust in the sense that it ensures boundedness of all signals in the face of bounded output disturbances. It is emphasized that the leakage modification (or any other modification) was not required in this case to establish boundedness.

On the other hand, a loss of robustness may occur if in addition to the output disturbance, one adds a bounded disturbance to the *regressor*  $x$ . In this scenario, the system might become excited over the parameter subspace which is not PE, leading to a loss of uniform asymptotic stability (u.a.s.) in the homogeneous part of (5.1)(5.2) (cf., [8], pg. 636). When u.a.s. is lost, counterexamples to stability exist with consequences of unbounded parameter drift (cf., [8], pg. 546). In this case it may be possible (although not yet proved) to recover exponential convergence if the regressor disturbance is in the form of a transient which decays sufficiently fast (e.g., exponentially). This case is common in practice due to transient effects in feedforward signals, and remains to be analyzed further.

## 6 Conclusions

The main results of the paper show that if the confluence matrix is positive definite the adaptive feedforward operator  $\mathcal{H}$  from the error  $e$  to the estimate  $\hat{y}$  is input-output equivalent to an adaptive system with a PE regressor. This implies that tracking error convergence is exponential for a large class of overparametrized adaptive feedforward systems. Explicit bounds on the convergence rate is given for purposes of systematic optimization and design.

The reader is warned, however, that the weakened PE condition does not enjoy all the robustness properties of full PE, and does not ensure parameter convergence over the entire space. It would also be prudent to restrict its application to adaptive feedforward rather than feedback applications, and especially applications where performance is judged by convergence of the tracking error rather than any requirement for convergence of the parameters,

## Acknowledgements

This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

## A APPENDIX A

**LEMMA 1** (*Sastry and Bodson [11]*) Consider the error equation,

$$e = \phi^T x \quad (\text{A.1})$$

$$\dot{\phi} = -\mu x x^T \phi \quad (\text{A.2})$$

where  $\phi(t), x(t) \in R^n$ . Let  $x$  be a bounded piecewise continuous function of  $t$  such that,

$$\|x(t)\| \leq \bar{x} < \infty; \text{ for all } t \geq 0 \quad (\text{A.3})$$

and let there exist constants  $\alpha_1, \alpha_2, \delta > 0$  such that the following PE condition is satisfied,

$$\alpha_1 I \leq \int_t^{t+\delta} x(t)x(t)^T dt \leq \alpha_2 I, \quad \text{for all } t \geq 0 \quad (\text{A.4})$$

Then the system (A.1)(A.2) is *globally exponentially stable*, i.e.,

$$\|x(t)\| \leq p_0 e^{-\alpha t} \quad (\text{A.5})$$

$$\|e\| \leq p_0 \bar{x} e^{-\alpha t} \quad (\text{A.6})$$

where,

$$\alpha \geq \frac{1}{2\delta} \ln \left( \frac{1}{1 - \alpha_3} \right) \quad (\text{A.7})$$

$$p_0 = \left( \frac{1}{1 - \alpha_3} \right)^{\frac{1}{2}} \|\phi(0)\| \quad (\text{A.8})$$

$$\alpha_3 = \frac{2\mu\alpha_1}{(1 + \mu\alpha_2\sqrt{n})^2} \quad (\text{A.9})$$

Letting  $\mu$  be sufficiently small (i.e., such that  $\mu < 1/(\alpha_2\sqrt{n})$ ), gives,

$$o_1 \simeq \mu\alpha_1/\delta \quad (\text{A.10})$$

$$p_0 \simeq (1 + \mu\alpha_1)\|\phi(0)\| \quad (\text{A.11})$$

**PROOF:** The proof follows directly from the development in Sastry and Bodson [11] pg. 73-75 (see in particular Theorem 2.5.3) specialized to the gradient adaptation algorithm (A.1)(A.2).

## References

- [1] K.J. Astrom and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records," In *Theory of Self-Adaptive Control Systems.*, pp. 96-111, Plenum Press, New York, 1966.
- [2] D.S. Bayard, J. Spanos and Z. Rahman, "A result on exponential tracking error convergence and persistent excitation," IEEE Trans. Auto. Contr. (to appear), 1998.
- [3] D.S. Bayard, *The role of the Confluence matrix in exponential tracking convergence of overparametrized adaptive feedforward systems.* Jet Propulsion Laboratory, Internal Document, JPL D-14442, March 15, 1997.
- [4] D.S. Bayard, "Necessary and sufficient conditions for LTI representations of adaptive systems with sinusoidal regressors," Proc. American Control Conference, Albuquerque NM, June 1997.
- [5] R.R. Bitmead, "Persistence of excitation conditions and the convergence of adaptive schemes," IEEE Trans. on Information Theory, vol. 30, pp. 183-191, March 1984.
- [6] R.R. Bitmead and B.D.O. Anderson, "Lyapunov techniques for the exponential stability of linear difference equations with random coefficients," IEEE Trans. Automatic Control, vol. AC-25, pp. 782-787, August 1980.
- [7] J.R. Glover, Jr., "Adaptive noise canceling applied to sinusoidal interferences," IEEE Trans. Acoustics, Speech and Signal Processing, vol. 25, no. 6, December 1977.
- [8] P.A. Ioannou and J. Sun, *Robust Adaptive Control.* Prentice Hall, Upper Saddle River, New Jersey, 1996.



- [9] R. Johansson, "Global Lyapunov stability and exponential convergence of direct adaptive control," *Int. J. Control*, vol. 50, no. 3, pp. 859-869, 1989.
- [10] K.S.Narendra and A.M. Annaswamy, *Stable Adaptive Systems*. Prentice-Hall, Englewood Cliffs, New Jersey, 1989.
- [11] S. Sastry and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*. Prentice Hall, 1989.
- [12] A. Weiss and D. Mitra, "Digital adaptive filters: Conditions for convergence, rates of convergence, effects of noise and errors arising from implementation," *IEEE Trans. on Information Theory*, vol. IT-25, pp. 637-652, November 1979.
- [13] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1985.
- [14] J. S-C Yuan and W.M. Wonham, "Probing signals for model reference identification," *IEEE Trans. Automatic Control*, vol. 22, pp. 530-538, August 1977.